Adaptive wavelet estimation of a compound Poisson process

Céline Duval*

Abstract

We study the nonparametric estimation of the jump density of a compound Poisson process from the discrete observation of one trajectory over $[0, T]$. We consider the microscopic regime when the sampling rate $\Delta = \Delta_T \to 0$ as $T \to \infty$. We propose an adaptive wavelet threshold density estimator and study its performance for the L_p loss, $p \geq 1$, over Besov spaces. The main novelty is that we achieve minimax rates of convergence for sampling rates Δ_T that vanish with T at arbitrary polynomial rates. More precicely, our estimator attains minimax rates of convergence provided there exists a constant $K \geq 1$ such that the sampling rate Δ_T satisfies $T\Delta_T^{2K+2} \leq 1$. If this condition cannot be satisfied we still provide an upper bound for our estimator. The estimating procedure is based on the inversion of the compounding operator in the same spirit as Buchmann and Grübel (2003).

AMS 2000 subject classifications: 62G99, 62M99, 60G50.
Keywords: Compound Poisson process, Discretely observed random process, Decompounding, Wavelet density estimation.

*GIS-CREST and CNRS-UMR 8050, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France.

1 Introduction

1.1 Statistical setting

Let R be a standard homogeneous Poisson process with intensity ϑ in $(0, \infty)$, we define the compound Poisson process X as

$$X_t = \sum_{i=1}^{R_t} \xi_i, \qquad t \geq 0$$

where the $\big(\xi_i\big)$ are independent and identically distributed random variables and independent of the Poisson process R.

Assume that we have discrete observations of the process X over $[0, T]$ at times $i\Delta$ for some $\Delta > 0$

$$\big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1} \rfloor \Delta}\big). \tag{1}$$

We focus on the *microscopic regime*, namely

$$\Delta = \Delta_T \to 0 \qquad \text{as } \; T \to \infty$$

and work under the following assumption.

Assumption 1. *The law of the ξ_i has density f which is absolutely continuous with respect to the Lebesgue measure.*

We denote by $\mathcal{F}(\mathbb{R})$ the space of densities with respect to the Lebesgue measure supported by $\mathbb{R}$. We investigate the nonparametric estimation of the density f on a compact interval $\mathcal{D}$ included in $\mathbb{R}$ from the observations (1). To that end we use wavelet threshold density estimators and study their rate of convergence uniformly over Besov balls for the following loss function

$$\big(\mathbb{E}\big[\|\widehat{f} - f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}, \tag{2}$$

where $\widehat{f}$ is an estimator of f, $p \geq 1$ and

$$\|f\|_{L_p(\mathcal{D})} = \Big(\int_{\mathcal{D}} |f(x)|^p dx\Big)^{1/p}.$$

We also denote by $\|f\|_{L_p(\mathbb{R})}$ the usual L_p norm for $p \geq 1$

$$\|f\|_{L_p(\mathbb{R})} = \Big(\int_{\mathbb{R}} |f(x)|^p dx\Big)^{1/p}.$$

We do not assume the intensity ϑ to be known: it is a nuisance parameter.

By Assumption 1, on the event $\{X_{i\Delta} - X_{(i-1)\Delta} = 0\}$ no jump occurred between $(i-1)\Delta$ and $i\Delta$ and the increment $X_{i\Delta} - X_{(i-1)\Delta}$ gives no information on f. In the microscopic regime many increments are zero, therefore to estimate f we focus on the nonzero increments and denote by N_T their number over $[0, T]$. In that statistical context different difficulties arise. First the sample size N_T is random. Second on the event $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the increment $X_{i\Delta} - X_{(i-1)\Delta}$ is not necessarily a realisation of the density f. Indeed even if Δ is small there is always a positive probability that more than one jump occurred between $(i-1)\Delta$ and $i\Delta$. Conditional on $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the law of $X_{i\Delta} - X_{(i-1)\Delta}$ has density given by (see Proposition 1 in Section 2 below)

$$\mathbf{P}_\Delta[f](x) = \sum_{m=1}^{\infty} \mathbb{P}\big(R_\Delta = m \big| R_\Delta \neq 0\big) f^{\star m}(x), \qquad \text{for } x \in \mathbb{R}, \tag{3}$$

where $\star$ is the convolution product and $f^{\star m} = f \star \ldots \star f$, m times.

Adaptive estimators of the density f in that statistical context already exists. Under the condition $T\Delta_T \leq 1$ (or $T\Delta_T^2 \leq 1$ if f is smooth enough), they attain minimax rates of convergence over Sobolev spaces for the L_2 loss (see Bec and Lacour [1], Comte and Genon-Catalot [4, 6] and Figueroa-López [10]). In this paper we try to answer the following questions.

- **i)** Is it possible to construct an estimator of f when Δ_T decays slowly to 0, for instance when Δ_T vanishes polynomially slowly with T.
- **ii)** Is it possible to construct adaptive wavelet estimators that attain, over Besov spaces for the L_p loss defined in (2), the classical minimax rates of convergence of the experiment where we observe T independent realisations of f.

Without loss of generality, assuming T is an integer if we observe T independent realisations of a density f of regularity s measured with the L_π norm, $\pi > 0$, it is possible to achieve the minimax rates of convergence for the L_p loss –up to constants and logarithmic factors– which is of the form

$$T^{-\alpha(s,\pi,p)}$$

where $\alpha(s, \pi, p) \leq 1/2$ (see for instance Donoho *et al.* [7] and (16) hereafter). When the process X is continuously observed over $[0, T]$, we have R_T independent and identically distributed realisations of f. Moreover for T large enough,

R_T is of the order of T. That is why we want compare the performance of estimators of f in the regime $\Delta_T \to 0$ with the classical minimax rate we would have if X were continuously observed.

1.2 Our Results

We build our estimator of f using equation (3) and proceed in two steps. The first step is the computation of the inverse of the operator $f \to \mathbf{P}_\Delta[f]$. The inverse takes the form

$$\mathbf{P}_\Delta^{-1}[\nu] = \sum_{m\geq 1}^{\infty} a_m(\vartheta, \Delta_T)\nu^{\star m}, \qquad \nu \in \mathcal{F}(\mathbb{R})$$

where the $\big(a_m(\vartheta, \Delta_T)\big)$ are explicit (see Proposition 1 below). They depend on the intensity ϑ and can be estimated. We take advantage of

$$f \approx \mathbf{L}_{\Delta,K}\big[\mathbf{P}_\Delta[f]\big], \tag{4}$$

where $\mathbf{L}_{\Delta,K}$ is the Taylor expansion of order K in Δ of $\mathbf{P}_\Delta^{-1}$. It depends only on $\big(\mathbf{P}_\Delta[f]^{\star m}, m = 1, \ldots, K+1\big)$. That step can be referred as decoumpounding as introduced in Buchmann *et al.* [2].

The second step consists in estimating the densities $\mathbf{P}_\Delta[f]^{\star m}$, for $m = 1, \ldots, K+1$. For that we use the N_T nonzero increments which are independent and with density $\mathbf{P}_\Delta[f]$. The difficulty here is that N_T is random. In Theorem 1 we show that conditional on N_T wavelet threshold estimators of $\mathbf{P}_\Delta[f]^{\star m}$ attain a rate of convergence –up to logarithmic factors– in $N_T^{-\alpha(s,\pi,p)}$. For T large enough we prove (see Proposition 2 in Section 5) that N_T concentrates around a deterministic value of the order of T, giving an unconditional rate of convergence in $T^{-\alpha(s,\pi,p)}$. We inject those estimators into $\mathbf{L}_{\Delta,K}$, defined in (4), and obtain an estimator of f that we call *estimator corrected at order K.*

The study of the rate of convergence of the estimator corrected at order K requires to control two distinct error terms. A deterministic one due the first step which is the error made when approximating f by $\mathbf{L}_{\Delta,K}\big[\mathbf{P}_\Delta[f]\big]$ in (4). And a statistical one due to the replacement of the $\mathbf{P}_\Delta[f]^{\star m}$ by estimators in the second step. The deterministic error decreases when K increases, then the idea is to choose K sufficiently large for the deterministic error term to be negligible in front of the statistical one. We give in Theorem 1 an upper bound for the rate of convergence of the estimator corrected at order K which is in –up to constants and logarithmic factors–

$$\max\{T^{-\alpha(s,\pi,p)}, \Delta_T^{K+1}\}.$$

It decreases with K and if there exists K_0 such that

$$T\Delta_T^{2K_0+2} \leq 1, \tag{5}$$

since $\alpha(s,\pi,p) \leq 1/2$ the estimator corrected at order K_0 attains the minimax rates of convergence. It follows that for every Δ_T polynomially decreasing with T, it is possible to exhibit K_0 such that (5) is valid and the estimator corrected at order K_0 provides a positive answer to **i)** and **ii)**. If no K enables to verify condition (5), Theorem 1 provides an upper bound for the rate of convergence of the estimator corrected at order K, in that case the estimator still provide a positive answer to **i)**.

In the case of a compound Poisson processes, the results of the present paper generalise to some extend those of Bec and Lacour [1], Comte and Genon-Catalot [4, 6] and Figueroa-López [10]. This is discussed in further details in Section 4. In Section 2 we give the main results of the paper. We properly define wavelet functions and Besov spaces used for the estimation before having a complete construction of the estimator corrected at order K. Then we give an upper bound for its rate of convergence for the L_p loss defined in (2), $p \geq 1$, uniformly over Besov balls. A numerical example illustrates the behavior of the estimator corrected at order K in Section 3. Finally Section 5 is dedicated to the proofs.

The model of this paper is central in many application fields *e.g.* statistical physics (see Moharir [17]), biology (see Huelsenbeck *et al.* [13]), financial series or mathematical insurance (see Scalas [19]). It is well adapted to study phenomena where random independent events occur at random times. For instance, in insurance failure theory these events can model the claims that insurance companies have to pay to the subscribers. The insurer's surplus at a given time t can be modeled by the following process

$$K(t) = K_0 + kt - X_t,$$

where K_0 is the capital of the company at time 0, the second term is a deterministic trend corresponding to the average income received from the subscribers and X is a compound Poisson process modeling the insurance claims occurring at random times with random amount of money at stake. It is the Cramér-Lundberg model; see Embrechts *et al.* [8] or Scalas [19]. Compound Poisson processes can also model the changes of an asset price in finance; see Masoliver *et al.* [15].

2 Main results

2.1 Besov spaces and wavelet thresholding

To estimate the densities $\big(\mathbf{P}_\Delta[f]^{\star m}, m = 1, \ldots, K+1\big)$ we use wavelet threshold density estimators and study their performance uniformly over Besov balls. In this paragraph we reproduce some classical results on Besov spaces, wavelet bases and wavelet threshold estimators (see Cohen [3], Donoho *et al.* [7] or Kerkyacharian and Picard [14]) that we use in the next sections.

Wavelets and Besov spaces

We describe the smoothness of a function with Besov spaces on $\mathcal{D}$. We recall here some well documented results on Besov spaces and their connection to wavelet bases (see Cohen [3], Donoho *et al.* [7] or Kerkyacharian and Picard [14]). Let $\big(\psi_\lambda\big)_\lambda$ be a regular wavelet basis adapted to the domain $\mathcal{D}$. The multi-index λ concatenates the spatial index and the resolution level $j = |\lambda|$. Set $\Lambda_j := \{\lambda, |\lambda| = j\}$ and $\Lambda = \cup_{j\geq -1}\Lambda_j$, for f in $L_p(\mathbb{R})$ we have

$$f = \sum_{j\geq -1}\sum_{\lambda\in\Lambda_j}\langle f, \psi_\lambda\rangle\psi_\lambda, \tag{6}$$

where $j = -1$ incorporates the low frequency part of the decomposition and $\langle ., \rangle$ denotes the usual L_2 inner product. We define Besov spaces in term of wavelet coefficients, for $s > 0$ and $\pi \in (0, \infty]$ a function f belongs to the Besov space $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ if the norm

$$\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} := \sup_{j\geq -1} 2^{j(s+1/2-1/\pi)}\Big(\sum_{\lambda\in\Lambda_j}|\langle f, \psi_\lambda\rangle|^\pi\Big)^{1/\pi} \tag{7}$$

is finite, with usual modifications if $\pi = \infty$.

We need additional properties on the wavelet basis $\big(\psi_\lambda\big)_\lambda$, which are listed in the following assumption.

Assumption 2. *For $p \geq 1$,*

- *We have for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1}2^{|\lambda|(p/2-1)} \leq \|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{|\lambda|(p/2-1)}.$$

- *For some $\mathfrak{C} > 0$, $\sigma > 0$ and for all $s \leq \sigma$, $J \geq 0$, we have*

$$\big\|f - \sum_{j \leq J} \sum_{\lambda \in \Lambda_j} \langle f, \psi_\lambda \rangle \psi_\lambda \big\|_{L_p(\mathcal{D})} \leq \mathfrak{C} 2^{-Js} \|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}. \tag{8}$$

- *If $p \geq 1$, for some $\mathfrak{C} \geq 1$ and for any sequence of coefficients $\big(u_\lambda\big)_{\lambda \in \Lambda}$,*

$$\mathfrak{C}^{-1} \Big\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \Big\|_{L_p(\mathcal{D})} \leq \Big\| \Big(\sum_{\lambda \in \Lambda} |u_\lambda \psi_\lambda|^2 \Big)^{1/2} \Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C} \Big\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \Big\|_{L_p(\mathcal{D})}. \tag{9}$$

- *For any subset $\Lambda_0 \subset \Lambda$ and for some $\mathfrak{C} \geq 1$*

$$\mathfrak{C}^{-1} \sum_{\lambda \in \Lambda_0} \|\psi_\lambda\|^p_{L_p(\mathcal{D})} \leq \int_{\mathcal{D}} \Big(\sum_{\lambda \in \Lambda_0} |\psi_\lambda(x)|^2 \Big)^{p/2} \leq \mathfrak{C} \sum_{\lambda \in \Lambda_0} \|\psi_\lambda\|^p_{L_p(\mathcal{D})}. \tag{10}$$

Property (8) ensures that definition (7) of Besov spaces matches the definition in terms of linear approximation. Property (9) ensures that $\big(\psi_\lambda\big)_\lambda$ is an unconditional basis of L_p and (10) is a super-concentration inequality (see Kerkyacharian and Picard [14] p. 304 and p. 306).

Wavelet threshold estimator

Let (ϕ, ψ) be a pair of scaling function and mother wavelet that generate a basis $\big(\psi_\lambda\big)_\lambda$ satisfying Assumption 2 for some $\sigma > 0$. We rewrite (6)

$$f = \sum_{k \in \Lambda_0} \alpha_{0k} \phi_{0k} + \sum_{j \geq 1} \sum_{k \in \Lambda_j} \beta_{jk} \psi_{jk},$$

where $\phi_{0k}(\bullet) = \phi(\bullet - k)$ and $\psi_{jk}(\bullet) = 2^{j/2} \psi(2^j \bullet - k)$ and

$$\alpha_{0k} = \int \phi_{0k}(x) f(x) dx$$
$$\beta_{jk} = \int \psi_{jk}(x) f(x) dx.$$

For every $j \geq 0$, the set Λ_j has cardinality 2^j and incorporates boundary terms that we choose not to distinguish in the notation for simplicity. An estimator of a function f is obtained when replacing the (α_{0k}) and (β_{jk}) by estimated values. In the sequel we uses (γ_{jk}) to design either (α_{0k}) or (β_{jk}) and (g_{jk}) for the wavelet functions (ϕ_{0k}) or (ψ_{jk}).

We consider classical hard threshold estimators of the form

$$\widehat{f}(\bullet)=\sum_{k\in\Lambda_0}\widehat{\alpha_{0k}}\phi_{0k}(\bullet)+\sum_{j=1}^{J}\sum_{k\in\Lambda_j}\widehat{\beta_{jk}}\mathbb{1}_{\left\{|\widehat{\beta_{jk}}|\geq\eta\right\}}\psi_{jk}(\bullet),$$

where $\widehat{\alpha_{0k}}$ and $\widehat{\beta_{jk}}$ are estimators of α_{0k} and β_{jk}, J and η are respectively the resolution level and the threshold, possibly depending on the data. Thus to construct $\widehat{f}$ we have to specify estimators $(\widehat{\gamma_{jk}})$ of the (γ_{jk}) and the coefficients J and η.

2.2 Construction of the estimator

Assume that we have $\lfloor T\Delta^{-1}\rfloor$ discrete data at times $i\Delta$ for some $\Delta>0$ of the process X

$$\big(X_\Delta,\ldots,X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big).$$

Introduce the increments

$$\mathbf{D}^\Delta X_i=X_{i\Delta}-X_{(i-1)\Delta},\qquad \text{for } i=1,\ldots,\lfloor T\Delta^{-1}\rfloor,$$

where $X_0=0$. They are independent and identically distributed since X is a compound Poisson process. Define

$$S_1=\inf\left\{j,\mathbf{D}^\Delta X_j\neq 0\right\}\wedge T$$
$$S_i=\inf\left\{j>S_{i-1},\mathbf{D}^\Delta X_j\neq 0\right\}\wedge T\quad \text{for } i\geq 1,$$

where S_i is the random index of the ith jump and

$$N_T=\sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor}\mathbb{1}_{\{\mathbf{D}^\Delta X_i\neq 0\}}$$

the random number of nonzero increments observed over $[0,T]$. By Assumption 1, on the event $\{\mathbf{D}^\Delta X_i=0\}$, no jump occurred between $(i-1)\Delta$ and $i\Delta$. In the microscopic regime when $\Delta=\Delta_T\to 0$ as T goes to infinity many increments are null and convey no information on f, hence for the estimation of f we focus on the nonzero ones

$$\big(\mathbf{D}^\Delta X_{S_1},\ldots,\mathbf{D}^\Delta X_{S_{N_T}}\big).$$

Proposition 1. *The distribution of the increment $\mathbf{D}^\Delta X_{S_1}$ has density with respect to the Lebesgue measure given by*

$$\mathbf{P}_\Delta[f]=\sum_{m=1}^{\infty}p_m(\Delta)f^{\star m},$$

where

$$p_m(\Delta) = \mathbb{P}\big(R_\Delta = m | R_\Delta \neq 0\big) = \frac{1}{e^{\vartheta\Delta} - 1}\frac{\big(\vartheta\Delta\big)^m}{m!}.$$

Let Δ_0 be such that

$$\sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta_0\big)^{m-2}}{m!} \leq 1.$$

For $\Delta \leq \Delta_0$, we have that

$$1 - \vartheta\Delta \leq p_1(\Delta) \leq 1.$$

It is straightforward to verify that the nonlinear operator $\mathbf{P}_\Delta$ is a mapping from $\mathcal{F}(\mathbb{R})$ to itself. The observations $\big(\mathbf{D}^\Delta X_{S_i}\big)$ are realisations of the density $\mathbf{P}_\Delta[f]$ and by Proposition 1 the weight $p_1(\Delta) \to 1$ in the limit $\Delta = \Delta_T \to 0$. It follows that for Δ_T small enough most of the $\big(\mathbf{D}^\Delta X_{S_i}\big)$ have distribution f. Then a naive method to estimate f is to apply classical density estimators to the $\big(\mathbf{D}^\Delta X_{S_i}\big)$. That estimator requires a convergence condition on Δ_T to achieve minimax rate of convergence (see Theorem 1). However we wish to construct an estimator that attains minimax rates of convergence with weaker conditions on Δ_T.

We adopt the estimating strategy of section 1.2 and construct an approximation of f.

Lemma 1. *The inverse $\mathbf{P}_\Delta^{-1}$ of $\mathbf{P}_\Delta$, such that for all densities f in $\mathcal{F}(\mathbb{R})$ if $\mathbf{P}_\Delta[f] = \nu$ we have $\mathbf{P}_\Delta^{-1}[\nu] = f$, is given by*

$$\mathbf{P}_\Delta^{-1}[\nu] = \frac{1}{\vartheta\Delta}\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m}(e^{\vartheta\Delta} - 1)^m \nu^{\star m}.$$

To build the estimator corrected at order K we use that $\mathbf{P}_\Delta^{-1}$ is a power series whose coefficients are equivalent to increasing powers of Δ. Then $\mathbf{L}_{\Delta,K}$ the Taylor expansion of order K in Δ of $\mathbf{P}_\Delta^{-1}$ is obtained by keeping the first $K+1$ terms of the inverse

$$\mathbf{L}_{\Delta,K}[\nu] = \frac{1}{\vartheta\Delta}\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m}(e^{\vartheta\Delta} - 1)^m \nu^{\star m}, \quad \nu \in \mathcal{F}(\mathbb{R}). \tag{11}$$

Next we construct wavelet threshold density estimators of the first $K+1$ convolution powers of $\mathbf{P}_\Delta[f]$ that will be plugged in (11). Define

$$\widehat{\gamma}_{jk}^{(m)} = \frac{1}{N_{T,m}}\sum_{i=1}^{N_{T,m}} g_{jk}\Big(\mathbf{D}_m^\Delta X_{S_i}\Big) \quad m \geq 1, \tag{12}$$

where $N_{T,m} = \lfloor N_T/m \rfloor \geq 1$ for large enough T and

$$\mathbf{D}_m^{\Delta} X_{S_i} = \mathbf{D}^{\Delta} X_{S_i} + \mathbf{D}^{\Delta} X_{S_{N_{T,m}+i}} + \cdots + \mathbf{D}^{\Delta} X_{S_{(m-1)N_{T,m}+i}}.$$

The $\left(\mathbf{D}^{\Delta} X_{S_i}\right)$ are independent and identically distributed with density $\mathbf{P}_{\Delta}[f]$, thus the $\left(\mathbf{D}_m^{\Delta} X_{S_i}\right)$ are independent and identically distributed with density $\mathbf{P}_{\Delta}[f]^{\star m}$. Let $\eta > 0$ and $J \in \mathbb{N} \setminus \{0\}$, define $\widehat{P_{\Delta,m}}$ the estimator of $\mathbf{P}_{\Delta}[f]^{\star m}$ over $\mathcal{D}$

$$\widehat{P_{\Delta,m}}(x) = \sum_k \widehat{\alpha}_{0k}^{(m)} \phi_{0k}(x) + \sum_{j=0}^{J} \sum_k \widehat{\beta}_{jk}^{(m)} \mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{(m)}| \geq \eta\right\}} \psi_{jk}(x), \quad x \in \mathcal{D}. \tag{13}$$

Definition 1. *We define* $\widetilde{f}_{T,\Delta}^{K}$ *the estimator corrected at order* K *for* K *in* $\mathbb{N}$ *and* x *in* $\mathcal{D}$ *as*

$$\widetilde{f}_{T,\Delta}^{K}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\left(e^{\widehat{\vartheta}_T \Delta} - 1\right)^m}{\widehat{\vartheta}_T \Delta} \widehat{P_{\Delta,m}}(x), \tag{14}$$

where

$$\widehat{\vartheta}_T = -\frac{1}{\Delta} \log(1 - \widehat{p}_T) \tag{15}$$

and

$$\widehat{p}_T = \frac{N_T}{\lfloor T\Delta^{-1} \rfloor}$$

is the empirical estimator of $p(\Delta) = \mathbb{P}(R_{\Delta} = 0) = 1 - e^{-\vartheta\Delta}$.

Lemma 1 justifies the form of the estimator corrected at order K.

2.3 Convergence rates

We estimate densities f which verify a smoothness property in term of Besov balls

$$\mathcal{F}(s, \pi, \mathfrak{M}) = \left\{f \in \mathcal{F}(\mathbb{R}), \|f\|_{\mathcal{B}_{\pi\infty}^{s}(\mathcal{D})} \leq \mathfrak{M}\right\},$$

where $\mathfrak{M}$ is a positive constant. We are interested in estimating f on the compact interval $\mathcal{D}$, that is why we only impose that its restriction to $\mathcal{D}$ belongs to a Besov ball.

Theorem 1. *We work under Assumptions 1 and 2. Let $\sigma > s > 1/\pi$, $p \geq 1 \wedge \pi$ and $\widehat{P_{\Delta_T,m}}$ be the threshold wavelet estimator of $\mathbf{P}_{\Delta_T}[f]^{\star m}$ on $\mathcal{D}$ constructed from (ϕ, ψ) and defined in (13). Take J such that*

$$2^J N_T^{-1} \log\big(N_T^{1/2}\big) \leq 1,$$

and

$$\eta = \kappa N_T^{-1/2} \sqrt{\log\big(N_T^{1/2}\big)},$$

for some $\kappa > 0$. Let

$$\alpha(s,p,\pi) = \min\Big\{\frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2\big(s+1/2-1/\pi\big)}\Big\}. \tag{16}$$

1) *The estimator $\widehat{P_{\Delta_T,m}}$ verifies for large enough T and sufficiently large $\kappa > 0$*

$$\sup_{\mathbf{P}_{\Delta_T}[f]^{\star m} \in \mathcal{F}(s,\pi,\mathfrak{M})} \Big(\mathbb{E}\big[\big\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_{L_p(\mathcal{D})}^p \big| N_T\big]\Big)^{1/p} \leq \mathfrak{C} N_T^{-\alpha(s,p,\pi)},$$

up to logarithmic factors in T and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi$.

2) *The estimator corrected at order K $\widetilde{f}_{T,\Delta_T}^K$ defined in (14) verifies for T large enough and any positive constants $\underline{\mathfrak{T}}$ and $\overline{\mathfrak{T}}$*

$$\sup_{\vartheta \in [\underline{\mathfrak{T}}, \overline{\mathfrak{T}}]} \sup_{f \in \mathcal{F}(s,\pi,\mathfrak{M})} \Big(\mathbb{E}\big[\big\|\widetilde{f}_{T,\Delta_T}^K - f\big\|_{L_p(\mathcal{D})}^p\big]\Big)^{1/p} \leq \mathfrak{C} \max\big\{T^{-\alpha(s,p,\pi)}, \Delta_T^{K+1}\big\},$$

up to logarithmic factors in T and where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{T}}, \overline{\mathfrak{T}}, K$.

The proof of Theorem 1 is postponed to Section 5. From a practical point of view when one computes the estimator $\widetilde{f}_{T,\Delta_T}^K$ from (1) the sample size is N_T, which is why in Theorem 1 we give the resolution level J and the threshold η as functions of N_T instead of replacing N_T by its deterministic counterpart. Explicit bound for κ is given in Lemma 4 hereafter.

In practice the values T and Δ_T are imposed or chosen by the practitioner. Theorem 1 ensures that the estimator corrected at order K attains the minimax rate $T^{-\alpha(s,p,\pi)}$ for the smallest K such that

$$\Delta_T = O\big(T^{-\frac{\alpha(s,p,\pi)}{K+1}}\big).$$

Since $\alpha(s,p,\pi) \leq 1/2$ it is sufficient to choose K such that

$$T\Delta_T^{2K+2} = O(1).$$

If Δ_T decays as a power of T *i.e.* if there exists $\delta > 0$ such that for some $\mathfrak{C} > 0$

$$\Delta_T \leq \mathfrak{C}T^{-\delta},$$

it is always possible to find a correction level K satisfying the previous constraint. The case $K = 0$ corresponds to the uncorrected estimator; it is the naive estimator one would compute making the approximation $f \approx \mathbf{P}_\Delta[f]$. In that case we get a rate of convergence in

$$\max\{T^{-\alpha(s,p,\pi)}, \Delta_T\},$$

which attains the minimax rate if $T^{\alpha(s,p,\pi)}\Delta_T \leq 1$. Since $\alpha(s,\pi,p) \leq 1/2$, it follows that the condition $T^{\alpha(s,p,\pi)}\Delta_T \leq 1$ already improves on the condition $T\Delta_T^2 \leq 1$ of Bec and Lacour [1], Comte and Genon-Catalot [4, 6] or Figueroa-López [10] (see Section 4 for comparison with other works).

3 A numerical example

We illustrate the behaviour of the estimator corrected at order K when K increases and compare its performance with an oracle: the wavelet estimator we would compute in the idealised framework where all the jumps are observed

$$\widehat{f}^{Oracle}(x) = \sum_k \widehat{\alpha}_{0k}^{Oracle}\phi_{0k}(x) + \sum_{j=0}^{J}\sum_k \widehat{\beta}_{jk}^{Oracle}\mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{Oracle}|\geq\eta\right\}}\psi_{jk}(x),$$

where

$$\widehat{\alpha}_{0k}^{Oracle} = \frac{1}{R_T}\sum_{i=1}^{R_T}\phi_{0k}(\xi_i) \quad \text{and} \quad \widehat{\beta}_{jk}^{Oracle} = \frac{1}{R_T}\sum_{i=1}^{R_T}\phi_{0k}(\xi_i),$$

R_T being the value of the Poisson process R at time T and (ξ_i) the jumps. The parameters J and η as well as the wavelet bases (ϕ,ψ) are the same as those used to compute the estimator corrected at order K. We consider a compound Poisson process of intensity $\vartheta = 1$ on $[0,T]$ and of compound law

$$f(x) = (1-a)f_1(x) + af_2(x)$$

where f_1 is the density of a Gaussian $\mathcal{N}(0,1)$ and f_2 of a Laplace with location parameter 1 and scale parameter 0.1, we take $a = 0.05$.

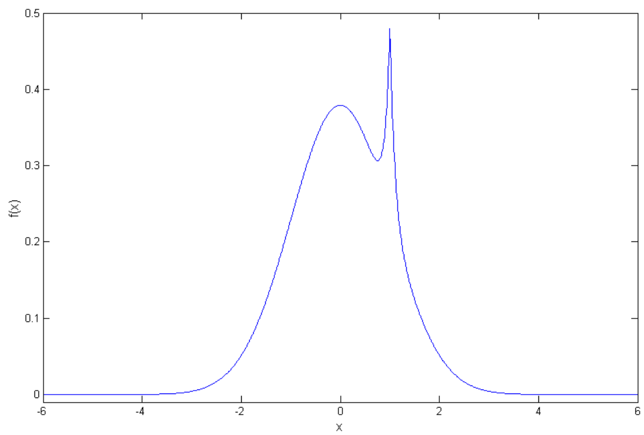

Figure 1: Density f : $f(x) = 0.95 f_1(x) + 0.05 f_2(x)$ $x \in [-6, 6]$.

We estimate the mixture f (see Figure 1) on $\mathcal{D} = [-6, 6]$ with the estimator corrected at order K for different values of K and study the results with the L_2 error. We also compare them with the oracle $\widehat{f}^{Oracle}$. Wavelet estimators are based on the evaluation of the first wavelet coefficients, to perform those we use Symlets 4 wavelet functions and a resolution level $J = 10$. Moreover we transform the data in an equispaced signal on a grid of length 2^L with $L = 8$, it is the binning procedure (see Härdle *et al.* [11] Chap. 12). The threshold is chosen as in Theorem 1. The estimators we obtain take the form of a vector giving the estimated values of the density f on the uniform grid $[-6, 6]$ with mesh 0.01. We use the wavelet toolbox of Matlab.

Figure 2 represents the corrected estimator for $K = 0$ and $K = 1$ and the oracle. All the estimators are evaluated on the same trajectory. They manage to reproduce the shape of the density f. As expected the oracle looks better than the other two and the uncorrected ($K = 0$) seems to make larger errors than the 1-corrected in estimating f. Figure 3 represents for every values in $[-6, 6]$ the absolute distance between those estimators –evaluated on the same trajectory– and the true density f. Therefore it enables to determine in which area an estimator fails to estimate f and to get an idea of the error made. The graphic was obtained after $M = 1000$ Monte-Carlo simulations of each estimator and averaging the results. The uncorrected estimator is not as good

as the estimator corrected at order 1. The oracle and the estimator corrected at order 1 seem to have similar performances. Each of the estimators makes larger errors around 1 which is where the density f is peaked.

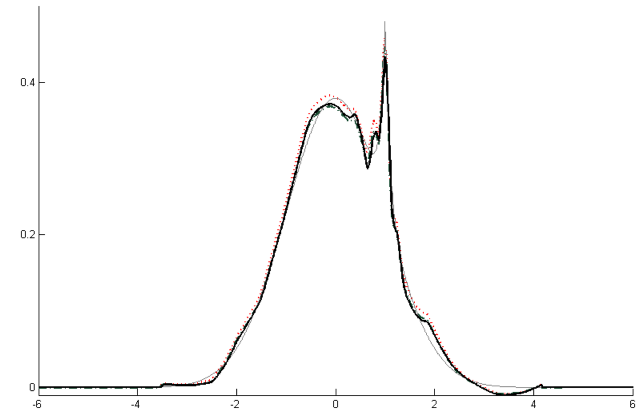

Figure 2: Estimators of the density f (plain grey) for $T = 10000$ and $\Delta = 0.1$: the uncorrected (dotted red), the 1-corrected (dashed green) and the oracle (plain dark).

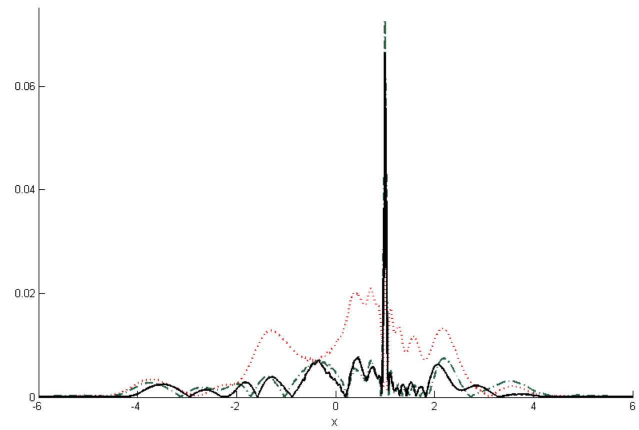

Figure 3: Mean absolute error between the estimators and the true density (M=1000, $T = 10000$ and $\Delta = 0.1$): the uncorrected (dotted red), the 1-corrected (dashed green) and the oracle (plain dark).

Evaluation of the L_2 errors enables to confirm the former graphical observation. We approximate the L_2 errors by Monte Carlo. For that we compute $M = 1000$ times each estimator (for $T = 10000$ and $\Delta = 0.1$) and approximate the L_2 loss by

$$\frac{1}{M}\sum_{i=1}^{M}\Big(\sum_{p=0}^{1200}\big(\widehat{f}(-6+0.01p)-f(-6+0.01p)\big)^2\times 0.01\Big),$$

where $\widehat{f}$ is one of the estimators. For each Monte Carlo iteration the corrected and oracle estimators are evaluated on the same trajectory. The results are reproduced in the following table.

Estimator	Oracle	$K=0$	$K=1$	$K=2$	$K=3$
L_2 error ($\times 10^{-4}$)	0.1117	0.1842	0.1353	0.1350	0.1350
Standard deviation ($\times 10^{-5}$)	0.3495	0.4434	0.4363	0.4366	0.4366

This confirms that there is an actual gain in considering the estimator corrected at order 1 instead of the uncorrected one. In the following table we estimate the $\big(p_m(\Delta)\big)$ defined in Proposition 1.

Estimated quantity	$\widehat{p_1}$	$\widehat{p_2}$	$\widehat{p_3}$
Estimation	0.9508	0.0476	0.0016
Standard deviation	0.0022	0.0022	0.0004

It turns out that without the correction we estimate the density f on a data set where 5% of the observations are realisations of a law which is not f. This explains why it is relevant to take them into account when estimating f. Considering more than 1 or 2 corrections is unnecessary as the L_2 losses get stable afterwards. The L_2 loss of the oracle is strictly lower than the loss of the estimator corrected at order K, even for large K. That difference is explained by the fact that to estimate the mth convolution power we do not use N_T data points but $N_{T,m} = \lfloor N_T/m \rfloor$. Therefore we do not loose in terms of rate of convergence, but we surely deteriorate the constants in comparison with the oracle. Numerical results are consistent with the theoretical results of Theorem 1 where we proved a rate of convergence for the estimator corrected at order K in

$$\max\left\{T^{-\alpha(s,p,\pi)}, \Delta_T^{K+1}\right\}.$$

Since $\alpha(s,p,\pi) \leq 1/2$, the rate decreases with K and becomes stable once $\Delta_T^{2K+2}T \leq \mathfrak{C}$. In the numerical example we took $T = 10000$ and $\Delta = 0.1$ thus $T\Delta^4 = 1$ which explains why in the example we did not observe improvements when correcting with K greater than 2.

4 Discussion

4.1 Relation to other works

A compound Poisson process is a pure jump Lévy process and can be studied accordingly using Lévy-Kintchine formula. Estimating the jump density f is then equivalent to estimating the Lévy measure since for compound Poisson process it is the product $\vartheta f(x)dx$. A possible estimation strategy in that case is to provide an estimator of the Fourier transform of the density. That strategy is quite different from the one introduced in this paper but is usually adopted when estimating the compound law of a compound Poisson process (see Figueroa-López [10], Comte and Genon-Catalot [4, 6] or Bec and Lacour [1]).

The nonparametric estimation of the Lévy measure from the discrete observation of a pure jump Lévy process from high frequency data (which corresponds to our microscopic regime $\Delta_T \to 0$) has been studied in great detail by Comte and Genon-Catalot [4, 6] and Figueroa-López [10]. In [10] the nonparametric

estimation of the Lévy density is made via a sieve estimator. They show that it attains minimax rates of convergence for the L_2 loss uniformly over a class of Besov functions for a sampling size Δ_T such that –with our notation– $T\Delta_T \leq 1$. Comte and Genon-Catalot [4, 6] construct an adaptive nonparametric estimator of the Lévy measure, which attains minimax rates of convergence on Sobolev spaces for the L_2 loss for a sampling size Δ_T such that $T\Delta_T \leq 1$ (or $T\Delta_T^2 \leq 1$ under smoother assumptions). Bec and Lacour [1] obtained similar results when $T\Delta_T^2 \leq 1$. The statistical setting of [6] is more general since they estimate the Lévy measure from observations of a Lévy process with a Brownian component.

Our result is limited to the Poisson case contrary to Bec and Lacour [1], Comte and Genon-Catalot [4] and Figueroa-López [10] who worked on the larger class of pure jump Lévy processes. However in the case of a Poisson process we generalise them since we provide an adaptive density estimator which attains minimax rates of convergence, for the L_p loss, $p \geq 1$, uniformly over Besov balls for regime where Δ_T is polynomially slow. If Δ_T decays even slower, for instance logarithmically in T, we still have an upper bound for the rate of convergence of our estimator.

4.2 Possible extensions

In this paper we give an adaptive minimax procedure for the estimation of the compound density of a compound Poisson process in the microscopic regime. The same estimation problem in an intermediate regime, namely when the process is observed at a sampling rate $\Delta > 0$ fixed, has been studied in van Es *et al.* [20] and in the more general setting of Lévy processes by Comte and Genon-Catalot [5] and Reiß [18]. van Es *et al.* [20] provide a consistent kernel density estimator of the compound density of a compound Poisson process of known intensity. They also focus on the nonzero increments for the estimation, but sidestep the problem of the random number of data N_T by assuming that they have a sample of a given size.

The estimator corrected at order K presented here should extend to intermediate regime where $\Delta_T \to \Delta_\infty < 1$ and the rate of convergence given in Theorem 1 should generalise in

$$\max\left\{T^{-\alpha(s,p,\pi)}, \Delta_\infty^{K+1}\right\}.$$

An improvement of the results would be the estimation of the compound density of renewal reward processes, or Continuous Time Random Walk, where it is no longer imposed that the elapsed time between jumps is exponentially distributed. Then the Lévy property is lost, the increments of the renewal process

are no longer independent nor identically distributed. An estimation strategy based on the Lévy-Kintchine formula is not possible. Such processes enable to model random phenomena where the elapse time between events is not memoryless; they have many applications for instance in finance (see Meerschaert *et al.* [16]), in biology (see Fedotov *et al.* [9]) or for modelling earthquakes (see Helmstetter *et al.* [12]).

5 Proof of Theorem 1

In the sequel $\mathfrak{C}$ denotes a generic constant which may vary from line to line. Its dependencies may be indicated in the index.

5.1 Proof of part 1) of Theorem 1

Preliminary lemmas

To prove part 1) of Theorem 1 we apply the general results of Kerkyacharian and Picard [14]. For that we establish some technical lemmas.

Lemma 2. *If f belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$ then for $m \geq 1$, $\mathbf{P}_\Delta[f]^{\star m}$ also belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$.*

Proof of Lemma 2. It is straightforward to derive $\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{L_1(\mathbb{R})} = 1$. The remainder of the proof is a consequence of the following result: Let $f \in \mathcal{B}^s_{\pi\infty}$ and $g \in L_1$ we have

$$\|f \star g\|_{s\pi\infty} \leq \|f\|_{s\pi\infty}\|g\|_{L_1(\mathbb{R})}. \tag{$\diamondsuit$}$$

To prove the ($\diamondsuit$) we use the following norm which is equivalent to the Besov norm (see [11])

$$\|\nu\|_{s\pi\infty} = \|\nu\|_{L_\pi(\mathbb{R})} + \|\nu^{(n)}\|_{L_\pi(\mathbb{R})} + \left\|\frac{w^2_\pi(\nu^{(n)},t)}{t^a}\right\|_\infty \tag{17}$$

where $s = n + a$, $n \in \mathbb{N}$ and $a \in (0,1]$, and w is the modulus of continuity

$$w^2_\pi(\nu,t) = \sup_{|h|\leq t}\left\|\mathbf{D}^h\mathbf{D}^h[\nu]\right\|_{L_\pi(\mathbb{R})},$$

where $\mathbf{D}^h[\nu](x) = \nu(x-h) - \nu(x)$. The result is a consequence of Young's inequality and elementary properties of the convolution product. We use the

definition (17) of the norm and treat each term separately. First Young's inequality gives

$$\|f_1 \star f_2\|_{L_\pi(\mathbb{R})} \leq \|f_1\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}. \tag{18}$$

Then the differentiation property of the convolution product leads for $n \geq 1$ to

$$\Big\|\frac{d^n}{dx^n}(f_1 \star f_2)\Big\|_{L_\pi(\mathbb{R})} = \Big\|\Big(\frac{d^n}{dx^n}f_1\Big) \star f_2\Big\|_{L_\pi(\mathbb{R})} \leq \Big\|\frac{d^n}{dx^n}f_1\Big\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}. \tag{19}$$

Finally translation invariance of the convolution product enables to get

$$\begin{aligned}\big\|\mathbf{D}^h\mathbf{D}^h[(f_1 \star f_2)^{(n)}]\big\|_{L_\pi(\mathbb{R})} &= \big\|(\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}]) \star f_2\big\|_{L_\pi(\mathbb{R})} \\ &\leq \big\|\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}]\big\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}.\end{aligned} \tag{20}$$

Inequality ($\diamondsuit$) is then obtained by bounding (17) with (18), (19) and (20) lead to the result. To complete the proof of Lemma 2, we apply $m-1$ times ($\diamondsuit$) which leads to

$$\forall m \in \mathbb{N}\setminus\{0\}, \qquad \big\|\mathbf{P}_\Delta[f]^{\star m}\big\|_{s\pi\infty} \leq \big\|\mathbf{P}_\Delta[f]\big\|_{s\pi\infty}.$$

The triangle inequality gives $\|\mathbf{P}_\Delta[f]^{\star m}\|_{s\pi\infty} \leq \|f\|_{s\pi\infty} \leq \mathfrak{M}$ which concludes the proof. $\square$

Lemma 3. *Let $2^j \leq N_T$ then for all $m \in \mathbb{N}\setminus\{0\}$ and for $p \geq 1$ we have*

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p\big|N_T\big] \leq \mathfrak{C}_{p,m,\|g\|_{L_p(\mathbb{R})},\mathfrak{M},\vartheta}N_T^{-p/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (12) *and*

$$\gamma_{jk}^{(m)} = \int g_{jk}(y)\mathbf{P}_\Delta[f]^{\star m}(y)dy. \tag{21}$$

Proof of Lemma 3. The proof is obtained with Rosenthal's inequality: let $p \geq 1$ and let $(Y_1,\ldots,Y_n)$ be independent random variables such that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\big[|Y_i|^p\big] < \infty$. Then there exists $\mathfrak{C}_p$ such that

$$\mathbb{E}\Big[\Big|\sum_{i=1}^n Y_i\Big|^p\Big] \leq \mathfrak{C}_p\Big\{\sum_{i=1}^n \mathbb{E}\big[|Y_i|^p\big] + \Big(\sum_{i=1}^n \mathbb{E}\big[|Y_i|^2\big]\Big)^{p/2}\Big\}. \tag{22}$$

The $\big(\mathbf{D}_m^{\Delta_T}X_{S_i}\big)$ are independent and identically distributed with common density $\mathbf{P}_{\Delta_T}[f]^{\star m}$ and $\mathbb{E}[\widehat{\gamma}_{jk}^{(m)}] = \gamma_{jk}^{(m)}$. Then $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is a sum of $N_{T,m} = \lfloor N_T/m \rfloor$

centered, independent and identically distributed random variables. It follows that

$$\begin{aligned}\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] &\le 2^p 2^{jp/2}\int |g(2^j y-k)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}(y)dy\\ &= 2^p 2^{j(p/2-1)}\int |g(z)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}\Big(\frac{z+k}{2^j}\Big)dz\\ &\le 2^p 2^{j(p/2-1)}\|g\|_{L_p(\mathbb{R})}^p \big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_\infty,\end{aligned}$$

where we made the substitution $z=2^j x-k$. To control $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty$ we use the Sobolev embeddings (see [3, 7, 11])

$$\mathcal{B}^s_{\pi\infty}\hookrightarrow \mathcal{B}^{s'}_{p\infty} \quad \text{and} \quad \mathcal{B}^{s'}_{\pi\infty}\hookrightarrow \mathcal{B}^{s}_{\infty\infty}, \tag{23}$$

where $p>\pi$, $s\pi>1$ and $s'=s-1/\pi+1/p$, it follows that

$$\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \le \mathfrak{C}_{s,\pi}\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{\mathcal{B}^s_{\pi\infty(\mathcal{D})}}.$$

We deduce from Lemma 2 that $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \le \mathfrak{C}_{s,\pi}\mathfrak{M}$. We get

$$\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] \le 2^p 2^{j(p/2-1)}\|g\|_{L_p(\mathbb{R})}^p\mathfrak{M}$$

and $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^2\big]\le\mathfrak{M}$ since $\|g\|_2^2=1$.

The accept-reject algorithm ensures that for all $n\ge 1$ the increments $\big(\mathbf{D}^{\Delta_T}X_{S_1},\dots,\mathbf{D}^{\Delta_T}X_{S_n}\big)$ are independent of N_T and then $N_{T,m}$. Indeed the $\big(\mathbf{D}^{\Delta_T}X_i, i=1,\dots,\lfloor T\Delta_T^{-1}\rfloor\big)$ are independent and identically distributed and the $\big(\mathbf{D}^{\Delta_T}X_{S_i}\big)$ are constructed with $S_i=\inf\big\{j>S_{i-1},\mathbf{D}^{\Delta_T}X_j\neq 0\big\}$. Therefore we can apply Rosenthal's inequality conditional on N_T to $\widehat{\gamma}_{jk}^{(m)}-\gamma_{jk}^{(m)}$ and derive for $p\ge 1$

$$\mathbb{E}\big[|\widehat{\gamma}_{jk}^{(m)}-\gamma_{jk}^{(m)}|^p\big|N_T\big]\le\mathfrak{C}_p\Big\{2^p\Big(\frac{2^j}{N_{T,m}}\Big)^{p/2-1}\|g\|^p_{L_p(\mathbb{R})}\mathfrak{M}+\mathfrak{M}^{p/2}\Big\}N_{T,m}^{-p/2}.$$

This concludes the proof. □

Lemma 4. *Choose j and c such that*

$$2^j N_T^{-1}\log\big(N_T^{1/2}\big)\le 1 \text{ and } c^2\ge\frac{16m}{3}\Big(\mathfrak{M}+\frac{c\|g\|_\infty}{6}\Big).$$

For all $m\in\mathbb{N}\setminus\{0\}$ and $r\ge 1$ let $\kappa_r=cr$. We have

$$\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)}-\gamma_{jk}^{(m)}|\ge\frac{\kappa_r}{2}N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}\Big|N_T\Big)\le N_T^{-r/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (12) and $\gamma_{jk}^{(m)}$ in (21).

Proof of Lemma 4. The proof is obtained with Bernstein's inequality. Consider $Y_1, \dots, Y_n$ independent random variables such that $|Y_i| \leq \mathfrak{A}$, $\mathbb{E}[Y_i] = 0$ and $b_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2]$. Then for any $\lambda > 0$,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n Y_i\Big| > \lambda\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2(b_n^2 + \frac{\lambda\mathfrak{A}}{3})}\Big). \tag{24}$$

For all $m \geq 1$, $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is a sum of $N_{T,m} = \lfloor N_T/m \rfloor$ centered independent and identically distributed random variables bounded by $2^{j/2}\|g\|_\infty$ and $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^2\big] \leq \mathfrak{M}$. The accept-reject algorithm ensures that for all $n \geq 1$ the increments $\big(\mathbf{D}^{\Delta_T} X_{S_1}, \dots, \mathbf{D}^{\Delta_T} X_{S_n}\big)$ are independent of N_T (see proof of Lemma 3), we apply Bernstein's inequality conditional on N_T. We have

$$\begin{aligned}
\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}\Big| N_T\Big)& \\
\leq & 2\exp\left(-\frac{\kappa_r^2 N_T^{-1}\log\big(N_T^{1/2}\big) N_{T,m}^2}{8\Big(N_{T,m}\mathfrak{M} + \frac{\kappa_r N_{T,m} N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}2^{j/2}\|g\|_\infty}{6}\Big)}\right) \\
= & 2\exp\left(-\frac{c^2 r N_T^{-1} N_{T,m}}{8\Big(\mathfrak{M} + \frac{\kappa_r N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}2^{j/2}\|g\|_\infty}{6}\Big)} r\log\big(N_T^{1/2}\big)\right).
\end{aligned}$$

Using that

$$m N_T^{-1} N_{T,m} = \frac{m}{N_T}\Big\lfloor \frac{N_T}{m} \Big\rfloor \geq \frac{3}{2},$$

for T large enough and $2^{j/2} N_T^{-1}\sqrt{\log\big(N_T^{1/2}\big)} \leq 1$ it follows that

$$\begin{aligned}
\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}\Big| N_T\Big)& \\
\leq 2\exp\left(-\frac{3c^2 r}{16m\big(\mathfrak{M} + \frac{\kappa_r\|g\|_\infty}{6}\big)} r\log\big(N_T^{1/2}\big)\right)&.
\end{aligned}$$

With $c^2 \geq \frac{16m}{3}\big(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\big)$ we get

$$\mathbb{P}\Big(|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\big(N_T^{1/2}\big)}\Big| N_T\Big) \leq N_T^{-r/2}.$$

The proof is complete. $\square$

Completion of proof of part 1) of Theorem 1

Part 1) of Theorem 1 is a consequence of Lemma 2, 3, 4 and of the general theory of wavelet threshold estimators of [14]. It suffices to have conditions (5.1) and (5.2) of Theorem 5.1 of [14], which are satisfied –Lemma 3 and 4– with $c(T) = N_T^{-1/2}$ and $\Lambda_n = c(T)^{-1}$ (with the notations of [14]). We can now apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [14] to obtain the result.

5.2 Proof of part 2) of Theorem 1

Preliminary result

The result of part 1) of Theorem 1 where given conditional on N_T. To prove part 2) we replace N_T by its deterministic counterpart. We introduce the following result.

Proposition 2. *For all $r > 0$, there exist $1 \leq \mathfrak{C}_\vartheta < \infty$, where $\vartheta \to \mathfrak{C}_\vartheta$ is continuous, such that*

$$1/\mathfrak{C}_\vartheta T^{-r} \leq \mathbb{E}\big[N_T^{-r}\big] \leq \mathfrak{C}_\vartheta T^{-r}.$$

Proof of Proposition 2. We have

$$N_T = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}},$$

where

$$\mathbb{E}\big[\mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}}\big] = p(\Delta_T) = 1 - \exp(-\vartheta\Delta_T).$$

Introduce $Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}} - p(\Delta_T)$, the Y_i are centered independent and identically distributed random variables bounded by 2 and $\mathbb{E}[Y_i^2] \leq p(\Delta_T)$, it follows from Bernstein's inequality (24) that for $\lambda > 0$

$$\mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| > \lambda\Big) \leq \exp\bigg(-\frac{\lfloor T\Delta_T^{-1}\rfloor\lambda^2}{2\big(p(\Delta_T) + \frac{2\lambda}{3}\big)}\bigg). \tag{25}$$

We choose $\lambda = p(\Delta_T)/2$, on the set $\big\{\big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\big| \leq \lambda\big\}$ we have

$$\lfloor T\Delta_T^{-1}\rfloor\frac{p(\Delta_T)}{2} \leq N_T \leq \lfloor T\Delta_T^{-1}\rfloor\frac{3p(\Delta_T)}{2}.$$

Moreover for Δ_T small enough we have that

$$\frac{\vartheta}{2} \leq p(\Delta_T) = 1 - \exp(-\vartheta\Delta_T) \leq \vartheta\Delta_T.$$

We have for all $\lambda > 0$

$$\mathbb{E}\big[N_T^{-r}\big] = \mathbb{E}\Big[N_T^{-r}\mathbb{1}_{\{|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor}-p(\Delta_T)|>\lambda\}}\Big] + \mathbb{E}\Big[N_T^{-r}\mathbb{1}_{\{|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor}-p(\Delta_T)|\leq\lambda\}}\Big].$$

Since for $r > 0$ the function $x \to x^{-r}$ is decreasing and $N_T \geq 1$ we have using (25) the upper bound

$$\begin{aligned}
\mathbb{E}\big[N_T^{-r}\big] &\leq \mathbb{P}\Big(\Big|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\Big| > \frac{p(\Delta_T)}{2}\Big) + \Big(\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \\
&\leq \exp\left(-\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)^2}{8\big(p(\Delta_T)+\frac{p(\Delta_T)}{3}\big)}\right) + \Big(\frac{\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \\
&\leq \exp\left(-\frac{3\vartheta}{64}T\right) + \Big(\frac{T\vartheta}{4}\Big)^{-r}.
\end{aligned}$$

For the lower bound we have

$$\mathbb{E}\big[N_T^{-r}\big] \geq \Big(\frac{3\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\Big)^{-r} \geq \Big(\frac{3T\vartheta}{2}\Big)^{-r}.$$

Then there exists $1 \leq \mathfrak{C}_\vartheta < \infty$ with $\vartheta \to \mathfrak{C}_\vartheta$ continuous such that

$$1/\mathfrak{C}_\vartheta T^{-r} \leq \mathbb{E}\big[N_T^{-r}\big] \leq \mathfrak{C}_\vartheta T^{-r}.$$

The proof is now complete. $\square$

Completion of proof of part 2) of Theorem 1

To prove Theorem 1 we define the quantity for K in $\mathbb{N}$ and x in $\mathcal{D}$

$$\widehat{f}_{T,\Delta}^K(x) = \sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}\frac{\big(e^{\vartheta\Delta}-1\big)^m}{\vartheta\Delta}\widehat{P_{\Delta,m}}(x).$$

It is the estimator of f one would compute if ϑ were known. We decompose the L_p error as follows

$$\begin{aligned}
\big(\mathbb{E}\big[\|\widetilde{f}_{T,\Delta_T}^K - f\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \leq &\big(\mathbb{E}\big[\|\widetilde{f}_{T,\Delta_T}^K - \widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \\
&+ \big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K - f\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p},
\end{aligned}$$

and control each term separately.

First we control $\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}-f\|^p_{L_p(\mathcal{D})}\big]$, using the triangle inequality we get

$$\begin{aligned}
&\Big(\mathbb{E}\Big[\Big\|\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}\frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{\vartheta\Delta_T}\widehat{P_{\Delta_T,m}}-\mathbf{P}^{-1}_{\Delta_T}\big[\mathbf{P}_{\Delta_T}[f]\big]\Big\|^p_{L_p(\mathcal{D})}\Big]\Big)^{1/p}\\
&\leq\sum_{m=1}^{K+1}\frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{m\vartheta\Delta_T}\big(\mathbb{E}\big[\big\|\widehat{P_{\Delta_T,m}}-\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p} \qquad (26)\\
&\quad+\sum_{m=K+2}^{\infty}\frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{m\vartheta\Delta_T}\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathbb{R})}. \qquad (27)
\end{aligned}$$

To bound (26) we use part 1) of Theorem 1 in which the supremum is taken over the class $\{\mathbf{P}_{\Delta_T}[f]^{\star m}\in\mathcal{F}(s,\pi,\mathfrak{M})\}$. With the inclusion

$$\{\mathbf{P}_{\Delta_T}[f]^{\star m},f\in\mathcal{F}(s,\pi,\mathfrak{M})\}\subset\mathcal{F}(s,\pi,\mathfrak{M})$$

and Proposition 2 applied with $r=\alpha(s,p,\pi)p>0$, we deduce the upper bound for $m\geq 1$

$$\begin{aligned}
\mathbb{E}\big[\big\|\widehat{P_{\Delta_T,m}}-\mathbf{P}^{-1}_{\Delta_T}\big[\mathbf{P}_{\Delta_T}[f]\big\|^p_{L_p(\mathcal{D})}\big]&\leq\mathfrak{C}\mathbb{E}\big[N_T^{-\alpha(s,p,\pi)p}\big]\\
&\leq\mathfrak{C}T^{-\alpha(s,p,\pi)p}, \qquad (28)
\end{aligned}$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K,\vartheta)$. To bound (27) Young's inequality and $\big\|\mathbf{P}_{\Delta_T}[f]\big\|_{L_1(\mathbb{R})}=1$ enable to get

$$\big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_{L_p(\mathbb{R})}\leq\big\|\mathbf{P}_{\Delta_T}[f]\big\|_{L_p(\mathbb{R})}\quad\text{for } m\geq 1.$$

The triangle inequality leads to $\big\|\mathbf{P}_{\Delta_T}[f]\big\|_{L_p(\mathbb{R})}\leq\|f\|_{L_p(\mathbb{R})}$ and we use the Sobolev embeddings (23) to get $\|f\|_{L_p(\mathbb{R})}\leq\mathfrak{C}_{s,\pi,p}\mathfrak{M}$. We derive the upper bound

$$\begin{aligned}
&\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{\vartheta\Delta_T}\big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_{L_p(\mathbb{R})}\\
&\qquad\leq\|f\|_{L_p(\mathbb{R})}\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\big(e^{\vartheta\Delta_T}-1\big)^m}{\vartheta\Delta_T}\\
&\qquad\leq\mathfrak{C}_{K,\vartheta,\mathfrak{M}}{\Delta_T}^{K+1}. \qquad (29)
\end{aligned}$$

Thus from (28) and (29) we obtain

$$\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\big\|\widehat{f}^K_{T,\Delta_T}-f\big\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\leq\mathfrak{C}\max\big\{T^{-\alpha(s,p,\pi)},\Delta_T^{K+1}\big\},$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K,\vartheta)$. Since $\vartheta\to\mathfrak{C}$ is continuous we get for $p\geq 1$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\big\|\widehat{f}^K_{T,\Delta_T}-f\big\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\leq \mathfrak{C}_{K,\mathfrak{M}}\max\big\{T^{-\alpha(s,p,\pi)},\Delta_T^{K+1}\big\},$$

where $\mathfrak{C}$ depends on $(s,\pi,p,\mathfrak{M},\phi,\psi,K)$

We now control $\mathbb{E}\big[\|\widetilde{f}^K_{T,\Delta_T}-\widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]$ and use (15) to derive

$$\widetilde{f}^K_{T,\Delta_T}=\sum_{m=1}^{K+1}\frac{(-1)^m}{m}\frac{\big((1-\widehat{p}_T)^{-1}-1\big)^m}{\log(1-\widehat{p}_T)}\widehat{P_{\Delta_T,m}},$$

where $\widehat{P_{\Delta_T,m}}$ does not depend on ϑ (see (12)). Define

$$G_m(x)=\frac{((1-x)^{-1}-1)^m}{\log(1-x)}.$$

The triangle inequality leads to

$$\begin{aligned}\big(\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta})-&\widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\\&\leq\sum_{m=1}^{K+1}\big(\mathbb{E}\big[\|\big(G_m(\widehat{p}_T)-G_m(p(\Delta_T))\big)\widehat{P_{\Delta_T,m}}\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p},\end{aligned}$$

where $p(\Delta_T)$ verifies $p(\Delta_T)=1-e^{-\vartheta\Delta_T}\leq\mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$ since

$$0<1-e^{-\underline{\mathfrak{T}}\Delta_T}\leq 1-e^{-\vartheta\Delta_T}\leq 1-e^{-\overline{\mathfrak{T}}\Delta_T}<1.$$

Moreover, we have

$$G_m'(x)=\frac{mx^{m-1}}{(1-x)^{m+1}\log(1-x)}+\frac{x^m}{(1-x)^{m+1}\big(\log(1-x)\big)^2},$$

then for all $m\geq 1$ $xG_m'(x)$ is continuous over $(0,1/2]$ and converges to 0 when $x\to 0$. We deduce

$$\begin{aligned}\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta})-&\widehat{f}^K_{T,\Delta_T}\|_{L_p(\mathcal{D})}\big]^{1/p}\\&\leq\mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}},K}\Delta_T^{-1}\mathbb{E}\big[\|\big(\widehat{p}_T-p(\Delta_T)\big)\widehat{P_{\Delta_T,m}}\|^p_{L_p(\mathcal{D})}\big]^{1/p}.\end{aligned}$$

Cauchy-Schwarz inequality leads to

$$\mathbb{E}\big[\big\|\big(\widehat{p}_T - p(\Delta_T)\big)\widehat{P_{\Delta_T,m}}\big\|_{L_p(\mathcal{D})}^p\big]^2 \\ \leq \mathbb{E}\Big[\big\|\big(\widehat{p}_T - p(\Delta_T)\big)\big\|_{2p}^{2p}\Big]\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|_{L_{2p}(\mathcal{D})}^{2p}\Big],$$

where using part 1) of Theorem 1 and that $N_T \geq 1$ we have

$$\begin{aligned}\mathbb{E}\Big[\big\|\widehat{P_{\Delta_T,m}}\big\|_{L_{2p}(\mathcal{D})}^{2p}\Big] &\leq \mathbb{E}\Big[\|\widehat{P_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_{2p}(\mathcal{D})}^{2p}\Big] + \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_{2p}(\mathcal{D})}^{2p} \\ &\leq \mathfrak{C}\mathbb{E}[N_T^{-2\alpha(s,p,\pi)p}] + \mathfrak{M}^{2p} \\ &\leq \mathfrak{C} \end{aligned} \tag{30}$$

where $\mathfrak{C}$ depends on $(s, \pi, p, \mathfrak{M}, \phi, \psi)$. We apply Rosenthal's inequality (22) to conclude the proof: $\widehat{p}_T - p(\Delta_T)$ is the sum of independent and identically distributed centered random variables

$$\big(Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T}X_i \neq 0\}} - p(\Delta_T), i \in \{1, \ldots, \lfloor T\Delta_T^{-1}\rfloor\}\big)$$

where $\mathbb{E}[|Y_i|^{2p}] \leq 2^{2p}\mathbb{E}\big[\mathbb{1}_{\{\mathbf{D}^{\Delta_T}X_i \neq 0\}}^{2p}\big] \leq \mathfrak{C}_{p,\mathfrak{T}}\Delta_T$ and $\mathbb{E}[|Y_i|^2] \leq \mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$. Rosenthal's inequality (22) gives

$$\mathbb{E}\big[\|\widehat{p}_T - p(\Delta_T)\|_{2p}^{2p}\big] \\ \leq \mathfrak{C}_{p,\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\lfloor T\Delta_T^{-1}\rfloor^{-2p}\big(\lfloor T\Delta_T^{-1}\rfloor\Delta_T + (\lfloor T\Delta_T^{-1}\rfloor\Delta_T)^p\big). \tag{31}$$

It follows from (30) and (31) that

$$\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K(\widehat{\vartheta}) - \widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}\big]^{1/p} \\ \leq \mathfrak{C}\Delta_T^{-1}\lfloor T\Delta_T^{-1}\rfloor^{-1}\big(T^{1/(2p)} + T^{1/2}\big),$$

where $\mathfrak{C}$ depends on $(s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{T}}, \overline{\mathfrak{T}}, K)$. We deduce for $p \geq 1$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\|\widehat{f}_{T,\Delta_T}^K(\widehat{\vartheta}) - \widehat{f}_{T,\Delta_T}^K\|_{L_p(\mathcal{D})}^p\big]\big)^{1/p} \\ \leq \mathfrak{C}\big(T^{-(1-1/(2p))} + T^{-1/2}\big)$$

where $\mathfrak{C}$ depends on $(s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{T}}, \overline{\mathfrak{T}}, K)$ and which is negligible compared to $T^{-\alpha(s,p,\pi)}$ since $\alpha(s, p, \pi) \leq 1/2$. The proof of Theorem 1 is now complete.

6 Appendix

6.1 Proof of Proposition 1

Let $x \in \mathbb{R}$, we have by stationarity of the increments of the process X

$$\begin{aligned}\mathbb{P}(\mathbf{D}^{\Delta} X_{S_1} \leq x) &= \mathbb{P}(X_\Delta \leq x | X_\Delta \neq 0)\\ &= \sum_{m=0}^{\infty} \mathbb{P}(X_\Delta \leq x | R_\Delta = m, R_\Delta \neq 0)\mathbb{P}(R_\Delta = m)\\ &= \sum_{m=1}^{\infty} p_m(\Delta)\mathbb{P}(X_\Delta \leq x | R_\Delta = m)\end{aligned}$$

where $\mathbb{P}(X_\Delta \leq x | R_\Delta = m) = \int_{-\infty}^{x} f^{\star m}(y)dy$ for $m \geq 1$. It follows

$$\mathbb{P}(\mathbf{D}^{\Delta} X_{S_1} \leq x) = \int_{-\infty}^{x} \mathbf{P}_\Delta[f](y)dy.$$

Immediate computation give the expression of $p_m(\Delta)$. For the control of $p_1(\Delta)$ the assertion $p_1(\Delta) \leq 1$ is immediate since $p_1(\Delta)$ is a probability. Moreover we have

$$\exp\big(\vartheta\Delta\big) - 1 = \vartheta\Delta\Big(1 + \vartheta\Delta \sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta\big)^{m-2}}{m!}\Big),$$

where

$$g(\Delta) := \sum_{m=2}^{\infty} \frac{\big(\vartheta\Delta\big)^{m-2}}{m!} = \frac{1}{\big(\vartheta\Delta\big)^2}\big(\exp\big(\vartheta\Delta\big) - 1 - \vartheta\Delta\big) \longrightarrow \frac{1}{2} \quad \text{as } \Delta \to 0.$$

Since g is continuous, there exists $\Delta_0 > 0$ such that for all $\Delta \leq \Delta_0$ we have $g(\Delta) \leq 1$. It follows for $\Delta \leq \Delta_0$ that

$$p_1(\Delta) \geq \frac{1}{1 + \vartheta\Delta} \geq 1 - \vartheta\Delta.$$

6.2 Proof of Lemma 1

Let $\mathbf{F}[f]$ denote the Fourier transform of f and take h such that $h = \mathbf{P}_\Delta[f]$. Using the one-to-one mapping between densities and their Fourier transform we show the relation for the Fourier transforms. The linearity of the Fourier transform and the relation $\mathbf{F}[f \star g] = \mathbf{F}[f]\mathbf{F}[g]$ give

$$\mathbf{F}[h] = \mathbf{F}\big[\mathbf{P}_\Delta[f]\big] = \frac{1}{e^{\vartheta\Delta} - 1} \sum_{m=1}^{\infty} \frac{\big(\vartheta\Delta\big)^m}{m!} \mathbf{F}[f]^m = \frac{\big(\exp(\vartheta\Delta\mathbf{F}[f]) - 1\big)}{e^{\vartheta\Delta} - 1},$$

from which we deduce

$$\mathbf{F}[f] = \frac{\log\left(1+(e^{\vartheta\Delta}-1)\mathbf{F}[h]\right)}{\vartheta\Delta} = \sum_{m=1}^{\infty}\frac{(-1)^{m+1}}{m}\frac{(e^{\vartheta\Delta}-1)^m}{\vartheta\Delta}\mathbf{F}[h]^m$$

as $\left\|(e^{\vartheta\Delta}-1)\mathbf{F}[h]\right\|_\infty < \left\|e^{\vartheta\Delta}-1\right\|_\infty < 1$ holds for $\Delta \leq \log 2$. We take the inverse Fourier transform of the equality to obtain the result.

Acknowledgements

This work is a part of the author's Ph.D thesis under the supervision of Marc Hoffmann whom I would like to thanks for his valuable remarks on this paper. The author's research is supported by a PhD GIS Grant.

References

[1] M. Bec, C. Lacour, Adaptive kernel estimation of the Lévy density, Hal preprint 0058322 (2011).

[2] B. Buchmann, R. Grübel, Decompounding: an estimation problem for Poisson random sums, The Annals of Statistics 31 (2003) 1054–1074.

[3] A. Cohen, Numerical Analysis of wavelet methods, Elsevier, 2003.

[4] F. Comte, V. Genon-Catalot, Nonparametric estimation for pure jump Lévy processes based on high frequency data, Stochastic Process. Appl. 119 (2009) 4088–4123.

[5] F. Comte, V. Genon-Catalot, Nonparametric adaptive estimation for pure jump Lévy processes, Annales de l'I.H.P., Probability and Statistics 46 (2010) 595–617.

[6] F. Comte, V. and Genon-Catalot, Estimation for Lévy processes from high frequency data within a long time interval, The Annals of Statistics 39 (2011) 803–837.

[7] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, D, Density estimation by wavelet Thresholding, The Annals of Statistics 24 (1996) 508–539.

[8] P. Embrechts, C. Klüppelberg, M. Mikosch, Modelling Extremal Events, Springer, 1997.

[9] S. Fedotov, A. Iomin, Probabilistic approach to a proliferation and migration dichotomy in the tumor cell invasion, Arxiv preprint 0711.1304v2 (2008).

[10] J.E. Figueroa-López, C. Houdré, Risk bounds for the nonparametric estimation of Lévy processes, IMS Lecture Notes-Monogr. Ser. High dimensional probability 51 (2006) 96–116.

[11] W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, approximation, and statistical applications, Springer-Verlag, New York, 1998.

[12] A. Helmstetter, D. and Sornette, Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models, The American Physical Society (2002).

[13] J.P. Huelsenbeck, B. Larget, D. Swofford, A Compound Poisson Process for Relaxing the Molecular Clock. Genetics Society of America (2000).

[14] G. Kerkyacharian, D. Picard, Thresholding algorithms, maxisets and well-concentrated bases, Test 9 (2000) 283–344.

[15] J. Masoliver, M. Montero, J. Perelló, G.H. Weiss, Direct and inverse problems with some generalizations and extensions, Arxiv preprint 0308017v2 (2008).

[16] M.M. Meerschaert, E. Scalas, Coupled continuous time random walk in finance, Physica A 370 (2006) 114–118.

[17] P.S. Moharir, Estimation of the compounding distribution in the compound Poisson process model for earthquakes, Proc. Indian Acad. Sci. 101 (1992) 347–359.

[18] M. Neumann, M. Reiß, Nonparametric estimation for Lévy processes from low-frequency observations, Bernoulli 15 (2009) 223–248.

[19] E. Scalas, The application of continuous-time random walks in finance and economics, Physica A 362 (2006) 225–239.

[20] B. van Es, S. Gugushvili, P. Spreij, A kernel type nonparametric density estimator for decompounding, Bernoulli 13 (2007) 672–694.